ELSEVIER

Paleopolyploidy and gene duplication in soybean and other legumes

Randy C Shoemaker[1], Jessica Schlueter[1] and Jeff J Doyle[2]

Two of the most important observations from whole-genome sequences have been the high rate of gene birth and death and the prevalence of large-scale duplication events, including polyploidy. There is also a growing appreciation that polyploidy is more than the sum of the gene duplications it creates, in part because polyploidy duplicates the members of entire regulatory networks. Thus, it may be important to distinguish paralogs that are produced by individual gene duplications from the homoeologous sequences produced by (allo)polyploidy. This is not a simple task, for several reasons, including the chromosomally cryptic nature of many duplications and the variable rates of gene evolution. Recent progress has been made in understanding patterns of gene and genome duplication in the legume family, specifically in soybean.

Addresses
[1] USDA-ARS and Iowa State University, G401 Agronomy Hall, Ames, Iowa 50011, USA
[2] LH Bailey Hortorium, Department of Plant Biology, Cornell University, Ithaca, New York 14853, USA

Corresponding author: Shoemaker, Randy C (rcsshoe@iastate.edu)

Current Opinion in Plant Biology 2006, **9**:104–109

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Nevin D Young and Rendy C Shoemaker

Available online 2nd February 2006

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2006.01.007

Introduction: *Glycine* in the context of legume phylogeny and cytology

Legumes are one of the three largest families of flowering plants, with nearly 20 000 species representing tremendous morphological, ecological, and genetic diversity [1,2]. The monophyletic subfamily Papilionoideae comprises more than two-thirds of these species and includes nearly all of the economically important crop legumes. Most crop legumes belong to the two major sister lineages that diverged from a common ancestor around 50 million years ago (mya) [3]: the Hologalegina, including *Lotus*, *Medicago*, and *Pisum*, and the phaseoloid–millettioid clade, containing *Glycine*, *Phaseolus*, and *Vigna*.

Most papilionoids are considered to be cytological diploids, with $x = 7$ or 8 in Hologalegina and $x = 10$ or 11 in the phaseoloids [4]. *Glycine* is an exception to these low chromosome numbers, and as such is a rarity among legumes [4]. The genus is divided into two subgenera, one of which includes only soybean (*Glycine max*) and its wild progenitor (*Glycine soja*), both of which are annual Asian species with $2n = 40$. Members of the subgenus *Glycine* are perennials, with around 25 'diploid' ($2n = 38$ or 40) species confined to Australia and around nine allotetraploid ($2n = 78$–80) taxa [5]. Thus, all extant species of *Glycine* are the products of an ancient genome duplication event, with multiple neopolyploid speciation events having been superimposed on this paleopolyploid genome within the past 50 000 years. A review on legume phylogenetics is presented by Cronk *et al*. elsewhere in this issue.

Genetic maps and cytogenetics

Genome duplications complicate comparative genome analyses because rearrangements of the genome are a common process that can occur soon after a genome duplication event [6,7]. Consequently, inversions and translocations mask many evolutionary connections among regions, and duplicate segments of chromosomes can remain hidden until a whole-genome sequence becomes available. Despite these impediments, nearly a decade ago, an RFLP map was produced in soybean that detected many duplicated segments across soybean's 20 linkage groups, as might have been expected from chromosome numbers. Unexpectedly, this map also identified several nested duplications [8], suggesting that the $2n = 20$ progenitor(s) of soybean had already undergone an earlier large-scale duplication event. This hypothesis has been supported in recent fluorescent *in situ* hybridization (FISH) studies of soybean. For example, two bacterial artificial chromosome (BAC) clones that were genetically anchored to the ends of linkage group E not only identified linkage group E but also hybridized to duplicated regions on two different chromosomes [9]. Although not discussed at the time, the presence of this event raised the question of how many other $2n = 10$ or 11 phaseoloids, or indeed other papilionoids, might be fundamentally but cryptically polyploid. The soybean linkage map is mute on the subject of the age of either duplication event.

Linkage maps are available for other legumes, and could potentially provide evidence of a shared duplication. Synteny appears to be less well-conserved among the genomes of legumes than among those of grasses. Gene loss and rearrangement makes the detection of synteny difficult even when it does exist [10], and RFLP maps

often do not provide adequate resolution for this purpose. Still, macrosynteny among legumes has been reported numerous times [11,12••,13–15]. Microsynteny also has been observed between soybean and *Medicago* [16,17,18•]. Microsynteny between homoeologous regions within the soybean genome has been estimated to range between 46% and approximately 90% [17,19,20]. These studies were based primarily on BAC fingerprints and limited cross-hybridization. Increased genomic sequence from *Glycine*, *Medicago* and *Lotus* will ultimately provide a complete picture of their shared and divergent genome structures. Meanwhile, many new insights have been provided by consideration of gene pairs, primarily from the large collections of expressed sequence tags (ESTs) available for these three taxa.

ESTs, Ks, and hypotheses of polyploidy

In contrast to linkage maps, paralogous genes provide an estimate of the age of the duplication that formed them because they accumulate synonymous substitutions in a roughly clock-like fashion. Hence, the age of the duplication can be estimated from the Ks (substitutions per synonymous site) value. Duplication and deletion (i.e. the birth and death) of genes are ongoing genomic processes. Most duplicate gene copies are lost within a relatively short timescale, producing a characteristic decay curve when the number of gene pairs is plotted against Ks [21]. The signature of a large-scale duplication is the presence of large number of paralogous gene pairs that show similar levels of divergence from one another. This will form a peak against the background of the birth/death curve for simple duplications. Early studies of soybean gene families identified numerous examples of putatively homoeologous sequence pairs [22–25], but EST collections provide the large amount of genes that are needed to identify such peaks. Recently, two studies of ESTs using slightly different methods identified such Ks peaks in diverse plant species, including soybean and *Medicago truncatula* [26••,27••].

As expected, soybean possesses two peaks (Figure 1). The median Ks values of these peaks were estimated to have occurred 14 and 44 Mya using the rate calibration favored by Schlueter *et al.* [26••]. Blanc and Wolfe [27••] used a different calibration that suggested that the two duplications were more recent. The Schlueter *et al.* [26••] dates are used here because they agree more closely with the divergence dates for legumes [3]. *Medicago* also possesses two Ks peaks, a dispersed younger peak that suggests an accumulation of regional duplications and a more ancient peak at about 58 Mya. The time of divergence of *Medicago* and *Glycine* has been estimated at around 50 Mya on the basis of the calibration of chloroplast gene phylogenies with legume fossil evidence [3]. If the *Medicago* value (58 Mya) is correct, then the event took place in the ancestor of both *Glycine* and *Medicago*, making the *Glycine* Ks value too young. Alternatively, if independent polyploidy events occurred in *Glycine* and *Medicago*, then the *Medicago* event must be younger than 50 Mya. Either way, the three events are so close that, given the large variance of Ks values in duplicated genes [28], further testing was warranted. Phylogenetic trees were estimated for 39 *Glycine* genes for which three or four copies exist. Tree topologies for the majority of genes were consistent with a single large-scale duplication event in the ancestor of *Glycine* and *Medicago* (Table 1; [29••]). This conclusion was supported by Mudge *et al.* [18•]. The rate of synonymous substitution in *Medicago* thus appears to be ~25–30% greater than that in soybean, accounting for the different age estimates for a shared Ks peak.

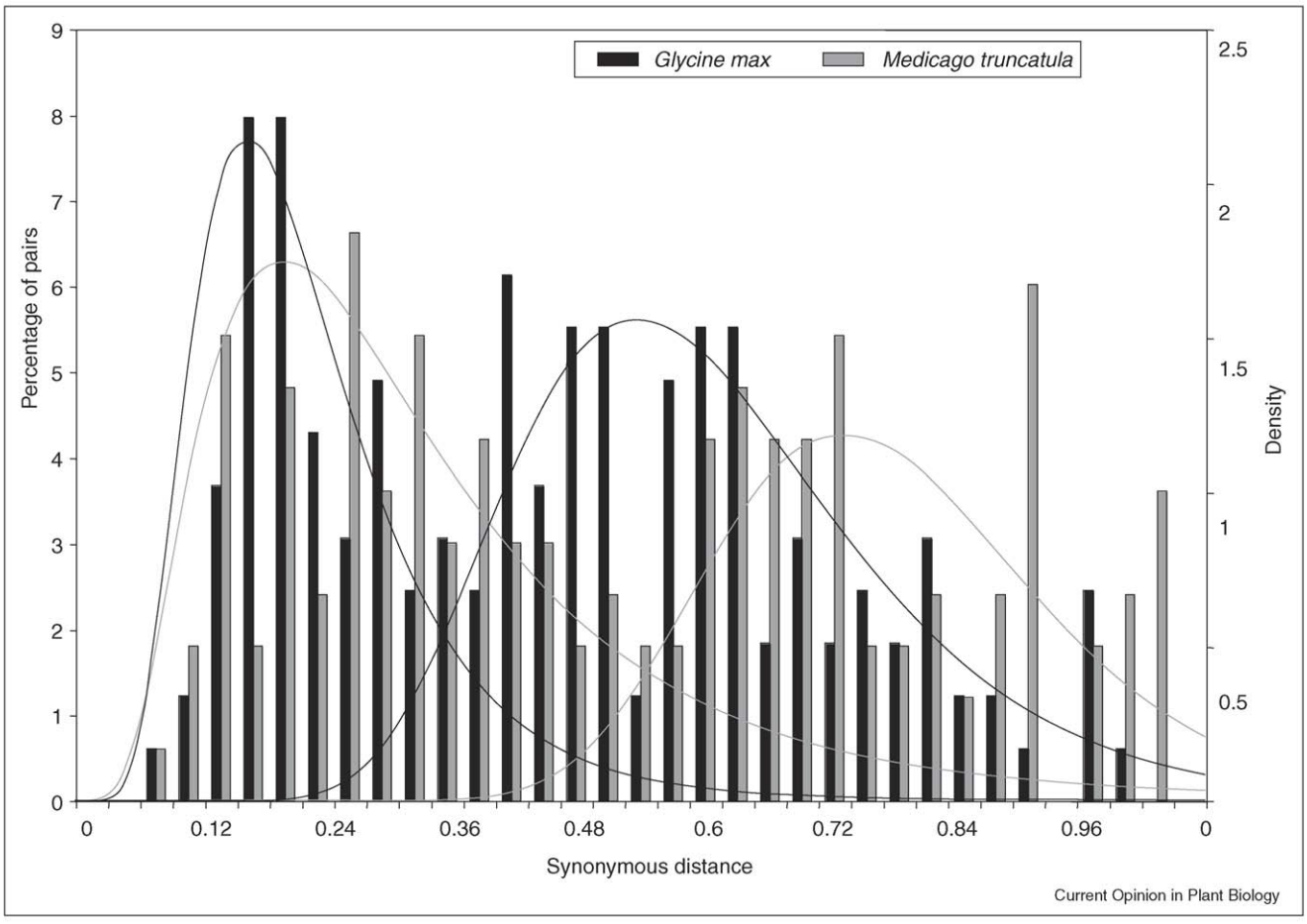
A major implication of this finding is that the Hologalegina plus the phaseoloid-millettiod group, which together comprise around 7000 species (a third of all legumes), are derived from an ancestor that experienced a large-scale gene duplication event. Moreover, given the very rapid radiation of legumes between 50–60 Mya [3], it is possible that the duplication took place in the ancestor of all papilionoids, or perhaps of all legumes.

Gene duplication, gene space, and legume evolution

Polyploidy has had a profound effect on the structure of the soybean genome. Hybridization-based studies in soybean suggest that the low-copy portion of the genome is present in approximately 2.6 copies [8]. Zhu *et al.* [30] estimated that about a quarter of duplicated genes have been lost since the last genome duplication event in soybean. An analysis of ESTs from the cultivar Williams 82 indicated that, on average, each gene family comprised 3.1 copies; this is fewer than would be expected if all of the copies from two rounds of whole-genome duplication were retained and expressed (R Shoemaker *et al.*, unpublished). Thus, about 25% of soybean gene duplicates have been silenced or lost.

The distribution of the remaining genes has become important as the community organizes for whole-genome sequencing projects, and the RFLP-based markers on the soybean genetic map are useful in estimating this. The soybean RFLP probes were generated with the restriction enzyme *Pst*I, a methylation-sensitive enzyme that cuts only in regions that are likely to be enriched for genes [31]. Using the distribution of BACs identified using RFLP probes and a pool of more than 110 000 BACs, a Poisson distribution of BAC 'hits' suggests that the gene space of soybean might be limited to as little as 24% of the genome [32••]. Although now thought to be an underestimate, the euchromatic region of *Medicago* was estimated to be only 20% of the genome [33], and this is expected to correspond with the gene space [34]. Although *Glycine* has a genome size almost twice that

Figure 1



Histograms of the percentage of duplicate gene pairs (primary y-axis) versus synonymous distance between pairs (x-axis) for *Glycine max* (black) and *Medicago truncatula* (grey). Overlying the histograms are Ln-normal distributions of the histograms. The density under each curve is shown on the secondary y-axis.

of *Medicago*, the gene density in *Glycine* and Medicago is similar (1 gene per 5.8–7.2 kb; [19]). Taken together, this suggests that the total gene number of soybean is likely to be twice that of *Medicago*.

Table 1

Phylogenetic resolution of shared duplication in *Glycine* and *Medicago*.

	Number of phylogenies	Percent of families
Hypothesis 1[a]	1	3%
Hypothesis 2[b]	22	56%
Neither[c]	4	10%
Equivocal[d]	5	13%
Unresolved[e]	7	18%

[a] *Glycine–Medicago* divergence before independent duplications.
[b] Shared *Glycine–Medicago* duplication before divergence.
[c] No support for either hypothesis.
[d] Null hypothesis, all duplications were independent.
[e] Conflicting resolution of phylogenies.

The distribution of duplicated genes in the gene space is of considerable interest. Expression shifts are expected among retained duplicates [21,35], and examples have been noted among soybean paralog pairs [36]. Non-coding sequences are thought to play a major role in the sub-functionalization of paralogous copies [37], whereas the coding regions of duplicated genes tend to evolve more slowly than those of singletons [38]. The evolutionary importance of duplication has long been appreciated [39] but, as noted above, polyploidy provides a new twist. When genes are duplicated in large numbers, there seems to be a bias in which genes are retained or lost. For example, genes that encode proteins that are involved in transcription or signal transduction are preferentially retained, whereas one copy of genes that are involved in DNA repair or defense is more likely to be silenced [40].

The identification of the genes that make legumes unique is of considerable interest. Genes that are

found only in legumes have been identified [41], but it is likely that many special morphological or ecological characteristics of legumes (e.g. symbiotic root nodulation) are shaped by gene families that have homologues in other flowering plants [42–46]. Gene duplication might therefore have played a significant role in the speciation of legumes. Duplications in the MADS-box [47] and *Cycloidea*-like [48–50] gene families have been suggested as factors in the evolution of the bilaterally symmetrical flowers that give papilionoid legumes their name. Another study supports the speciation origin of clusters of rapidly evolving apyrase genes in *Medicago*, *Glycine* and *Lotus*, and suggests that gene duplication occurred just before, or early in, the evolution of legumes [51].

Beyond determining more precisely the importance of duplications in key functional genes, it will be important to ascertain whether key duplications are correlated as part of large-scale or genome-wide events. One of the most intriguing questions in plant genome evolution is the degree to which polyploidy is responsible for major innovations that led to significant evolutionary radiations [52]. Timing the extent of the duplication event shared by *Medicago* and *Glycine* is an important step towards answering this question in legumes.

Conclusions and perspectives

The discovery of evidence for many gene duplications in the genomes of legumes is not surprising. Gene duplications are an accepted source of evolutionary novelty upon which natural or human selection can work. Duplicated genes are not only fundamentally interesting but are also of serious practical concern, complicating physical and genetic mapping, genome sequencing studies, reverse genetic approaches to understanding gene function, and plant adaptation [53].

Of particular fundamental interest to biologists are large-scale or whole-genome events. These seem to preserve duplicates to a greater degree than do single gene duplications, and also appear to retain certain classes of genes [40]. It has become apparent that polyploidy can generate substantial structural and epigenetic changes [30]. Key to understanding the overall structure of the genome is a set of markers that will identify duplicate regions [30]. Given the prevalence of large-scale genome duplications [27••,28], those markers should be compatible across species. Most crop legumes, we now know, are fundamentally polyploid, with more recent but still ancient polyploidy superimposed in *Glycine* and even more recent autopolyploidy in alfalfa. Revealing the contribution of gene and genome duplications to the evolution and domestication of these plants is a significant research program in itself, and has direct applications to future improvements in these crop species.

Acknowledgements

The authors gratefully acknowledge funding from the National Science Foundation (NSF), grants DED-0516673 and DBI-0321664 to JJD and DEB-0501877 to RCS, as well as funding from the United Soybean Board.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doyle JJ, Luckow MA: **The rest of the iceberg. Legume diversity and evolution in a phylogenetic context**. *Plant Physiol* 2003, **131**:900-910.

2. Lewis GP, Schrire BD, Mackinder BA, Lock M (Eds): *Legumes of the World*. Royal Botanic Garden; 2005.

3. Lavin M, Herendeen PS, Wojciechowski MF: **Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary**. *Syst Biol* 2005, **54**:575-594.

4. Goldblatt P: **Cytology and the phylogeny of leguminosae**. *Royal Botanic Gardens* 1981:427-463.

5. Doyle JJ, Doyle JL, Rauscher JT, Brown AHD: **Diploid and polyploid reticulate evolution throughout the history of perennial soybeans (*Glycine* subgenus *Glycine*)**. *New Phytol* 2004, **161**:121-132.

6. Ohno S: *Evolution by Gene Duplication.* Springer-Verlag; 1970.

7. Pickett FB, Meeks-Wagner DR: **Seeing double: appreciating genetic redundancy**. *Plant Cell* 1995, **7**:1347-1356.

8. Shoemaker RC, Polzin J, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP *et al.*: **Genome duplication in soybean (*Glycine* subgenus *soja*)**. *Genetics* 1996, **144**:329-338.

9. Pagel J, Walling JG, Young N, Shoemaker RC, Jackson SA: **Segmental duplications within the *Glycine max* genome revealed by fluorescence *in situ* hybridization of bacterial artificial chromosomes**. *Genome* 2004, **47**:764-768.

10. Ku HM, Vision T, Liu J, Tanksley SD: **Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny**. *Proc Natl Acad Sci USA* 2000, **97**:9121-9126.

11. Lee JM, Grant D, Vallejos CE, Shoemaker RC: **Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes**. *Theor Appl Genet* 2001, **103**:765-773.

12. •• Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB *et al.*: **Estimating genome conservation between crop and model legume species**. *Proc Natl Acad Sci USA* 2004, **101**:15289-15294.
The authors developed cross-species markers and genetically mapped putative orthologs to resolve macrosyntenic relationships among legumes. They created the first consensus comparative map for six legume species.

13. Kalo P, Seres A, Taylor SA, Jakab J, Keve Z, Kereszt A, Endre G, Ellis THN, Kiss GB: **Comparative mapping between *Medicago sativa* and *Pisum sativum***. *Mol Genet Genomics* 2004, **272**:235-246.

14. Boutin SR, Young ND, Olson TC, Yu ZH, Shoemaker RC, Vallejos CE: **Genome conservation among three legume genera detected with DNA markers**. *Genome* 1995, **38**:928-937.

15. Zhu H, Choi HK, Cook DR, Shoemaker RC: **Bridging model and crop legumes through comparative genomics**. *Plant Physiol* 2005, **137**:1189-1196.

16. Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND: **Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana***. *Genome* 2004, **47**:141-155.

17. Yan HH, Mudge J, Kim DJ, Larson D, Shoemaker RC, Cook DR, Young ND: **Estimates of conserved microsynteny among the genomes of *Glycine max, Medicago truncatula* and *Arabidopsis thaliana***. *Theor Appl Genet* 2003, **106**:1256-1265.

18. • Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, Young ND: **Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana***. *BMC Plant Biol* 2005, **5**:15.
This paper describes a region of exceptional synteny among legumes (*Glycine* and *Medicago*) and between legumes and *Arabidopsis thaliana*. The paper also clearly defines duplicated segments on different *Medicago* chromosomes.

19. Foster-Hartnett D, Mudge J, Larsen D, Danesh D, Yan H, Denny R, Penuela S, Young ND: **Comparative genomic analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group G**. *Genome* 2002, **45**:634-645.

20. Marek LF, Mudge J, Darnielle L, Grant D, Hanson N, Paz M, Huihuang Y, Denny R, Larson K, Foster-Hartnett D *et al.*: **Soybean genomic survey: BAC-end sequences near RFLP and SSR markers**. *Genome* 2001, **44**:572-581.

21. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**:1151-1155.

22. Lee JS, Verma DPS: **Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization**. *EMBO J* 1984, **3**:2745-2752.

23. Hightower R, Meagher R: **Divergence and differential expression of soybean actin genes**. *EMBO J* 1985, **4**:1-8.

24. Grandbastien MA, Berry-Lowe S, Shirley BW, Meagher R: **Two soybean ribulose-1, 5-bisphospahate carboxylase small subunit genes share extensive homology even in distant flanking sequences**. *Plant Mol Biol* 1986, **7**:451-465.

25. Nielsen NC, Dickinson C, Cho TJ, Thanh VH, Scallon BJ, Rischer RL, Sims TL, Drews GN, Goldberg RB: **Characterization of the glycinin gene family in soybean**. *Plant Cell* 1989, **1**:313-328.

26. •• Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC: **Mining EST databases to resolve evolutionary events in major crop species**. *Genome* 2004, **47**:868-876.
These authors used EST paralogs and synonymous distances of eight plant species to estimate coalescence times of the gene pairs. Like Blanc and Wolfe [27••], they concluded that *Glycine* and *Medicago* probably underwent large-scale genome duplication events. They propose that if the two genera shared a duplication event, then *Medicago* accumulated synonymous mutations at a rate 1.3 times faster than *Glycine*.

27. •• Blanc G, Wolfe KH: **Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes**. *Plant Cell* 2004, **16**:1667-1678.
The authors of this paper used ESTs identify pairs of paralogs in 14 species. Like Schlueter *et al*. [26••], they used synonymous substitution among the ESTs to estimate the relative ages of gene duplications. They concluded that *Glycine* and *Medicago* probably underwent large-scale duplication events.

28. Zhang L, Vision TJ, Gaut BS: **Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana***. *Mol Biol Evol* 2002, **19**:1464-1473.

29. •• Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ: **Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families**. *Syst Biol* 2005, **54**:441-454.
A phylogenetic analysis of 39 gene families in legumes (*Glycine* and *Medicago*), each with three or four members, is used to test whether gene duplications occurred separately in each lineage after divergence, or whether the different lineages share a round of duplications. The authors conclude that the evidence strongly supports a common duplication event. This means that approximately 7000 other legume species probably shared the same event.

30. Zhu T, Schupp JM, Oliphant A, Keim P: **Hypomethylated sequences: characterization of the duplicate soybean genome**. *Mol Gen Genet* 1994, **244**:638-645.

31. Grant D, Cregan P, Shoemaker RC: **Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis***. *Proc Natl Acad Sci USA* 2000, **97**:4168-4173.

32. •• Mudge J, Yan HH, Denny RL, Howe DK, Danesh D, Marek LF, Retzel E, Shoemaker RC, Young ND: **Soybean bacterial artificial chromosome contigs anchored with RFLPs: insights into genome duplication and gene clustering**. *Genome* 2004, **47**:361-372.
Using a combination of observed and expected RFLP sampling redundancy within a BAC library, the authors of this paper conclude that the gene space of soybean probably consisted of less than 275 Mb, or about 25% of the genome. This was the first report of estimated gene-space size in soybean.

33. Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook D, Huguet T, de Jong JH, Fransz PF, Bisseling T: **Integration of the FISH pachytene and genetic maps of *Medicago truncatula***. *Plant J* 2001, **27**:49-58.

34. Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O *et al.*: **A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa***. *Genetics* 2004, **166**:1463-1502.

35. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV: **Selection in the evolution of gene duplications**. *Genome Biol* 2002, **3**: RESEARCH0008.

36. Granger C, Coryell V, Khanna A, Keim P, Vodkin L, Shoemaker RC: **Identification, structure, and differential expression of members of a BURP domain containing protein family in soybean**. *Genome* 2002, **45**:693-701.

37. Lockton S, Gaut BS: **Plant conserved non-coding sequences and paralogue evolution**. *Trends Genet* 2005, **21**:60-65.

38. Jordan IK, Wolf YI, Koonin EV: **Duplicated genes evolve slower than singletons despite the initial rate increase**. *BMC Evol Biol* 2004, **4**:22-32.

39. Taylor JS, Raes J: **Duplication and divergence: the evolution of new genes and old ideas**. *Annu Rev Genet* 2004, **38**:615-643.

40. Blanc G, Wolfe KH: **Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution**. *Plant Cell* 2004, **16**:1679-1691.

41. Graham MA, Silverstein KA, Cannon SB, VandenBosch KA: **Computational identification and characterization of novel genes from legumes**. *Plant Physiol* 2004, **135**:1179-1197.

42. Schauser L, Wieloch W, Stougaard J: **Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus***. *J Mol Evol* 2005, **60**:229-237.

43. Geurts R, Fedorova E, Bisseling T: **Nod factor signaling genes and their function in the early stages of *Rhizobium* infection**. *Curr Opin Plant Biol* 2005, **8**:346-352.

44. Takane K, Tajima S, Kouchi H: **Two distinct uricase II (nodulin 35) genes are differentially expressed in soybean plants**. *Mol Plant Microbe Interact* 1997, **10**:735-741.

45. Frendo P, Jimenez MJH, Mathieu C, Duret L, Gallesi D, Van de Sype G, Herouart D, Puppo A: **A *Medicago truncatula* homoglutathione synthetase is derived from glutathione synthetase by gene duplication**. *Plant Physiol* 2001, **126**:1706-1715.

46. Rodriguez-Llorente ID, Perez-Hormaeche J, El Mounadi K, Dary M, Caviedes MA, Cosson V, Kondorosi A, Ratet P, Palomares AJ: **From pollen tubes to infection threads: recruitment of *Medicago* floral pectic genes for symbiosis**. *Plant J* 2004, **39**:587-598.

47. Dong Z.-c., Zhong Z, Liu C.-w., Luo J.-h., Yang J, Huang W.-h., Hu X.-h., Wang TL, Luo D: **Floral patterning in *Lotus japonicus***. *Plant Physiol* 2005, **137**:1272-1282.

48. Fukuda T, Yokoyama J, Maki M: **Molecular evolution of cycloidea-like genes in Fabaceae**. *J Mol Evol* 2003, **57**:588-597.

49. Citerne HL, Luo D, Pennington RT, Coen E, Cronk QCB: **A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the leguminosae**. *Plant Physiol* 2003, **131**:1042-1053.

50. Ree RH, Citerne HL, Lavin M, Cronk QC: **Heterogeneous selection on LEGCYC paralogs in relation to flower morphology and the phylogeny of *Lupinus* (Leguminosae)**. *Mol Biol Evol* 2004, **21**:321-331.

51. Cannon SB, McCombie WR, Sato S, Tabata S, Denny R, Palmer L, Katari M, Young ND, Stacey G: **Evolution and microsynteny of the apyrase gene family in three legume genomes**. *Mol Genet Genomics* 2003, **270**:347-361.

52. Adams KL, Wendel JF: **Polyploidy and genome evolution in plants**. *Curr Opin Plant Biol* 2005, **8**:135-141.

53. Paterson AH: **Polyploidy, evolutionary opportunity, and crop protection**. *Genetica* 2005, **123**:191-196.